NUMERICAL MATHEMATICS
A Journal of Chinese Universities (English Series)

Vol. **16** No. 1, pp. 92-96
February 2007

A Hybrid Method for Nonlinear Least Squares Problems

Zhongyi Liu[1,*] and Linping Sun[2]

[1] *School of Sciences, Hehai University, Nanjing 210098, China.*
[2] *Department of Mathematics, Nanjing University, Nanjing 210093, China.*

Received November 29, 2005; Accepted (in revised version) May 18, 2006

Abstract. A negative curvature method is applied to nonlinear least squares problems with indefinite Hessian approximation matrices. With the special structure of the method, a new switch is proposed to form a hybrid method. Numerical experiments show that this method is feasible and effective for zero-residual, small-residual and large-residual problems.

Key words: Nonlinear least squares; switch; hybrid method; negative curvature; BP factorization.

AMS subject classifications: 65K05, 90C30

1 Introduction

Consider nonlinear least squares problems

$$\min_{x\in R^n} F(x) = \frac{1}{2} f(x)^T f(x) = \frac{1}{2}\sum_{i=1}^{m} f_i(x)^2 \tag{1}$$

where $m \geq n$, $f : \mathbb{R}^n \to \mathbb{R}^m \in C^2(\Omega)$, $\Omega \in \mathbb{R}^n$ is an open convex set and $f_i(x)$ is the component function of $f(x)$. The gradient of $F(x)$ is

$$g(x) = J(x)^T f(x), \tag{2}$$

where $J(x)$ is the Jacobian matrix of $f(x)$, and the Hessian matrix is

$$G(x) = J(x)^T J(x) + \sum_{i=1}^{m} f_i(x)\nabla^2 f_i(x).$$

Set

$$M(x) = J(x)^T J(x), \quad W(x) = \sum_{i=1}^{m} f_i(x)\nabla^2 f_i(x). \tag{3}$$

*Correspondence to: Zhongyi Liu, School of Sciences, Hehai University, Nanjing 210098, China. Email: zhyi@hhu.edu.cn

Then

$$G(x) = M(x) + W(x). \tag{4}$$

Using the special structures of the object function $F(x)$ and the Hessian matrix $G(x)$, many effective methods have been developed. Among them a fundamental method is Gauss-Newton method which neglects the nonlinear term $W(x)$ in $G(x)$. In other words, a search direction is given by

$$J(x_k)^T J(x_k) p_k = -J(x_k)^T f(x_k). \tag{5}$$

The following theorem shows the convergence of the Gauss-Newton method.

Theorem 1.1. *Suppose that $F(x) \in C^2(\Omega)$, x^* is a local minimum of (1), $J(x)$ and $G(x)$ are Lipschitz continuous in Ω, and for all $x \in \Omega$, $J(x)$ is of full rank. If $\|J(x)\| \leq \delta$, $\|(J(x)^T J(x))^{-1}\| \leq \tau$, where δ and τ are constants, then Gauss-Newton iteration is well-defined for all $x \in \Omega$, and*

$$\|x^{(k+1)} - x^*\| \leq \|(J(x^*)^T J(x^*))^{-1} W(x^*)\| \|x^{(k)} - x^*\| + \mathcal{O}(\|x^{(k)} - x^*\|^2). \tag{6}$$

From the theorem above, whether Gauss-Newton method can succeed depends on whether the neglected term $W(x)$ is important, that is to say, whether $W(x)$ is a small part in $G(x)$. The Gauss-Newton method has quadratic rate of convergence for zero residual problems where $f(x^*) = 0$ or $W(x^*) = 0$.

The search direction can also be obtained by

$$(J(x^{(k)})^T J(x^{(k)}) + \lambda_k I) p^{(k)} = -J(x^{(k)})^T f(x^{(k)}) \tag{7}$$

where the nonnegative scalar λ_k is used to make $J(x^{(k)})^T J(x^{(k)}) + \lambda_k I$ positive definite. This formula is first proposed by Levenberg [4] and Marquardt [5], and is therefore called Levenberg-Marquardt method.

Another method takes advantage of $W(x)$ in $G(x)$, which is necessary for large residuals. One of this type of methods is due to Dennis-Gay-Welsh [6]. Since

$$\nabla^2 f_i(x^{(k+1)}) s^{(k)} = \nabla f_i(x^{(k+1)}) - \nabla f_i(x^{(k)}), \tag{8}$$

we have

$$f_i(x^{(k+1)}) \nabla^2 f_i(x^{(k+1)}) s^{(k)} = f_i(x^{(k+1)}) (J_{k+1} - J_k)^T e_i, \tag{9}$$

which leads to

$$\sum_{i=1}^{m} f_i(x^{(k+1)}) \nabla^2 f_i(x^{(k+1)}) s^{(k)} = (J_{k+1} - J_k)^T f^{(k+1)}. \tag{10}$$

Set $y^\sharp = (J_{k+1} - J_k)^T f^{(k+1)}$. Then W_{k+1} satisfies

$$W_{k+1} s = y^\sharp. \tag{11}$$

The Dennis-Gay-Welsh method gave the updating formula for W_k and scale strategy as follows:

$$W_{k+1} = \tau W_k + \frac{(y^\sharp - \tau W_k s) y^T + y (y^\sharp - \tau W_k s)^T}{y^T s} - \frac{(y^\sharp - \tau W_k s)^T s}{(y^T s)^2} y y^T, \tag{12}$$

where

$$y = J_{k+1}^T f^{(k+1)} - J_k^T f^{(k)}, \quad y^\sharp = J_{k+1}^T f^{(k+1)} - J_k^T f^{(k+1)}, \tag{13}$$

$$\tau = \min\left(1, \frac{|s^T y^\sharp|}{|s^T W_k s|}\right). \tag{14}$$

Since the nonlinear least squares problems are also general optimization problems, the hybrid methods using Gauss-Newton method or the Levenberg-Marquadt method can be used for zero residual problems, otherwise quasi-Newton method is adopted. Because hybrid methods can be adapted to a suitable method for many problems, hybrid methods are also called switch methods. Fletcher-Xu [3] proposed a switch strategy with little computation and superlinear convergence provided that

$$\frac{|F_{k-1} - F_k|}{F_{k-1}} < \rho, \tag{15}$$

where $\rho = 0.2$ is suggested. Fletcher-Xu's hybrid method is one of the most efficient algorithms for solving nonlinear least squares problems.

2 Algorithm

From Theorem 1.1, when $\|(J(x^*)^T J(x^*))^{-1} W(x^*)\| \leq \theta < 1$, the Gauss-Newton method converges at superlinear rate at least. As a result, the matrix norms $\|J(x^*)^T J(x^*)\|$ and $\|W(x^*)\|$ are compared in practice. It is quite expensive to compute the matrix norms directly. The algorithm to be proposed here will not require matrix norms and extra memory. And this algorithm is different from that of Fletcher-Xu, which is based on the comparison of function values in order to distinguish the residuals. Our proposed algorithm is based on Theorem 1.1 directly. Now we introduce the algorithm in detail. By (11),

$$\begin{aligned}
(J_{k+1}^T J_{k+1} + W_{k+1})s &= J_{k+1}^T J_{k+1} s + W_{k+1} s \\
&= J_{k+1}^T J_{k+1} s + (J_{k+1} - J_k)^T f^{(k+1)} = y, \\
J_{k+1}^T J_{k+1} s &= y - (J_{k+1} - J_k)^T f^{(k+1)} \\
&= J_{k+1}^T f^{(k+1)} - J_k^T f^{(k)} - J_{k+1}^T f^{(k+1)} + J_k^T f^{(k+1)} \\
&= J_k^T (f^{(k+1)} - f^{(k)}).
\end{aligned} \tag{16}$$

For convenience, we rewrite (11) as follows

$$W_{k+1} s = (J_{k+1} - J_k)^T f^{(k+1)}. \tag{17}$$

Note that $\|J_{k+1}^T J_{k+1} s\|$ and $\|W_{k+1} s\|$ can be regarded as the approximations of $\|J_{k+1}^T J_{k+1}\|$ and $\|W_{k+1}\|$ in the search direction s. Now we propose a new switch as follows

$$T_k : \qquad \|J_{k+1}^T J_{k+1} s\| \geq \gamma \|W_{k+1} s\|, \tag{18}$$

where $\gamma > 1$, and

$$B_k = \begin{cases} J_k^T J_k & \text{if } T_k \text{ is true}, \\ J_k^T J_k + W_k & \text{otherwise}. \end{cases} \tag{19}$$

Table 1: The numerator is our result, and the denominator is the result of [3].

Problems	Residual type (R)	N_{it}	N_f	N_g
Rosenbrock	Z	8/16	11/28	8/16
Wood		18/36	40/66	18/39
Box		5/5	5/7	5/6
Beale		6/7	7/20	6/11
Cragg		10/8	11/18	10/11
Watson6	S	5/5	5/7	5/6
Watson9		5/5	5/6	5/5
Watson12		5/5	5/8	5/7
Watson20		8/5	48/7	8/6
Osborne1		10/9	12/14	10/9
Osborne2		11/9	17/19	11/9
Bard		5/5	5/7	5/6
Madsen		4/5	5/8	4/5
Kowalik		6/6	9/16	6/8
Meyer	L	7/7	15/28	7/10
Freudenstein and Roth		9/6	19/10	9/6
Jennrich and Sampson		6/6	16/15	6/7

We now describe the proposed algorithm as follows:

Algorithm

Given x^0, W_0, $k = 0$, ε and evaluate J_0, f_0.

1. Evaluate d_k by $B_k d_k = -J_k^T f_k$, where B_k is given by (19).
2. Inexact line search
$$F(x_k + \alpha_k d_k) < F(x_k) + \rho\alpha_k g_k^T d_k.$$
3. Set $x_{k+1} = x_k + \alpha_k d_k$ and evaluate J_{k+1}, f_{k+1}.
4. Termination test
$$\|g(x_{k+1})\| < \varepsilon \text{ or } F_k - F_{k+1} <= 10^{-8} \max\{1, F_{k+1}\}.$$
5. If $\|y_k^T s_k\| > \eta\|y_k\|\|s_k\|$, update W_k by Dennis-Gay-Welsh updating formula and scale strategy, otherwise updating is skipped. Set $k := k + 1$, goto step 1.

Since positive definiteness of W_k cannot be assured by the Dennis-Gay-Welsh formula, we apply BP factorization method to the system $B_k d_k = -J_k^T f_k$, and a negative curvature direction is obtained [9].

3 Numerical results

The algorithm presented in this paper was implemented in a C++ code. This section discusses our computational results which will be compared with the results of Fletcher-Xu [3]. All testing was performed on a 1.1GHZ Pentium III PC with 256M memory. The test problems were obtained from Fletcher-Xu [3].

In Table 1, N_{it}, N_f, N_g represent the number of iterations, function evaluating, and gradient evaluating, respectively. The numerator is the result obtained from this algorithm, while the denominator is the result of Fletcher-Xu.

During inexact line search, we adopt backtracking line search with $\rho = 10^{-4}$. We also set $B_0 = I$, $\varepsilon = 10^{-8}$ for stopping test in step 4, and $\eta = 0.01$ in step 5. In the switch strategy, we set $\gamma = 100$.

It is observed that the proposed method is effective and feasible. However, for many test functions, numerical results are only a little better than that of Fletcher-Xu. Moreover, the convergence and stability have not been verified. This will be a future research topic.

References

[1] Brown K M, Dennis J E. A new algorithm for nonlinear least squares curve fitting, in Mathematical Software, Rice J R, ed. Academic Press , New York, 391-396, 1971.

[2] Toint P L. On large-scale nonlinear least squares calculatings. SIAM J. Sci. Stat. Comput., 1987, 8: 416-435.

[3] Fletcher R, Xu C. Hybrid methods for nonlinear least squares. IMA J. Numer. Anal., 1987, 7: 371-389.

[4] Levenberg K. A method for the solution of certain problems in least squares. Q. Appl. Math., 1944, 2: 164-168.

[5] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math., 1963, 11: 431-441.

[6] Dennis J E, Gay D M, Welsch R E. Algorithm 573-NL2SOL, An adaptive nonlinear least-squares algorithm. ACM T. Math. Software, 1981, 7: 348-368.

[7] Toint PH L. On large scale nonlinear least squares calculations. SIAM J Sci. Stat. Comput, 1987, 8: 416-435.

[8] Bunch J R, Kaufman L, Parlett B N. Decomposition of a symmetric matrix. Numer. Math., 1976, 27: 95-109.

[9] Zhang J Z, Xu C X. A Class of Trust Region Dogleg Methods for Unconstrained Minimization. Research Report MA-94-09, City University of Hong Kong, Hong Kong, 1994.

[10] Bunch J R, Parlett B N. Direct method for solving symmetric indefinite systems of linear equations. SIAM J. Numer. Anal., 1971, 8: 639-655.

[11] Nocedal J, Wright S J. Numerical Optimization. Springer, 1999.

[12] Dennis J E, Robert J R, Schnabel B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Inc, 1983.

[13] Mohamed J L, Walsh J. Numerical Algorithm. Oxford, 1986.